



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 103042**

**TO: Vanessa L Ford**  
**Location: cm1/8d17/8e12**  
**Art Unit: 1645**

*Sept 7*, 2003

**Case Serial Number: 10054354**

**From: P. Sheppard**  
**Location: CM1-1E03**  
**Phone: (703) 308-4499**

**sheppard@uspto.gov**

### **Search Notes**



103042

STIC-Biotech/ChemLib

From: Chan, Christina  
Sent: Thursday, September 04, 2003 8:43 AM  
To: Ford, Vanessa; STIC-Biotech/ChemLib  
Subject: RE: In Re:10/054354 Sequence Search

Please rush. Thanks Chris

Chris Chan

TC 1600 New Hire Training Coordinator and SPE 1644  
308-3973  
CM-1, 9B19

-----Original Message-----

From: Ford, Vanessa  
Sent: Wednesday, September 03, 2003 9:21 PM  
T : Chan, Christina  
Subject: In Re:10/054354 Sequence Search

Please search SEQ ID NO: 1

Please include interference searches. Please rush.

Vanessa L. Ford  
Biotechnology Patent Examiner  
Office: CM1 8D17  
Mailbox: CM1 8E12  
Phone: 703.308.4735

Point of Contact  
P. Sheppard  
Telephone number: (703) 308-4499

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: 9/7/03  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

CONFIDENTIAL

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2003, 14:00:54 ; Search time 96 Seconds

(without alignments)

53.761 Million cell updates/sec

Title: US-10-054-354-1

Perfect score: 110

Sequence: 1 KSTVGVEGLKHDWDGSPILK 20

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_muc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	278	2 Q9F472	Q9F472 ehrlichia c
2	110	100.0	278	2 Q9R8A8	Q9R8A8 ehrlichia c
3	110	100.0	278	2 Q9R8A7	Q9R8A7 ehrlichia c
4	110	100.0	278	2 Q9R3J3	Q9R3J3 ehrlichia c
5	110	100.0	278	2 Q9R8A6	Q9R8A6 ehrlichia c
6	110	100.0	278	2 Q9R8A9	Q9R8A9 ehrlichia c
7	110	100.0	278	2 Q9R8A5	Q9R8A5 ehrlichia c
8	110	100.0	307	2 Q9ZGJ1	Q9ZGJ1 ehrlichia c
9	79	71.8	280	2 Q9ADV3	Q9ADV3 ehrlichia c
10	79	71.8	280	2 Q9F473	Q9F473 ehrlichia c
11	76	69.1	280	2 Q52107	Q52107 ehrlichia c
12	73	66.4	246	2 Q9RH35	Q9RH35 ehrlichia c
13	73	66.4	275	2 Q93DD4	Q93DD4 ehrlichia c
14	73	66.4	275	2 Q93DD1	Q93DD1 ehrlichia c
15	73	66.4	276	2 Q9S817	Q9S817 ehrlichia c
16	73	66.4	276	2 Q8GGU0	Q8GGU0 ehrlichia c

17	73	66.4	280	2 Q93DD3	Q93DD3 ehrlichia c
18	73	66.4	280	2 Q9ZGM9	Q9ZGM9 ehrlichia c
19	73	66.4	280	2 Q8S816	Q8S816 ehrlichia c
20	72	65.5	281	2 Q93DD2	Q93DD2 ehrlichia c
21	72	65.5	281	2 Q9ACT9	Q9ACT9 ehrlichia c
22	70	63.6	288	2 Q9ZGJ2	Q9ZGJ2 ehrlichia c
23	63	57.3	290	2 Q9AEU3	Q9AEU3 cowdria rum
24	60	54.5	265	2 Q9AF99	Q9AF99 cowdria rum
25	60	54.5	270	2 Q9AF98	Q9AF98 cowdria rum
26	60	54.5	275	2 Q93B59	Q93B59 cowdria rum
27	60	54.5	276	2 Q93B60	Q93B60 cowdria rum
28	60	54.5	276	2 Q93B53	Q93B53 cowdria rum
29	60	54.5	277	2 Q93B65	Q93B65 cowdria rum
30	60	54.5	278	2 Q93B52	Q93B52 cowdria rum
31	60	54.5	278	2 Q93B57	Q93B57 cowdria rum
32	60	54.5	278	2 Q93B56	Q93B56 cowdria rum
33	60	54.5	287	2 Q93B29	Q93B29 cowdria rum
34	60	54.5	287	2 Q9R425	Q9R425 cowdria rum
35	60	54.5	287	2 Q46331	Q46331 cowdria rum
36	60	54.5	290	2 Q46324	Q46324 cowdria rum
37	60	54.5	290	2 Q93B64	Q93B64 cowdria rum
38	60	54.5	290	2 Q46333	Q46333 cowdria rum
39	60	54.5	290	2 Q46332	Q46332 cowdria rum
40	60	54.5	290	2 Q46330	Q46330 cowdria rum
41	58	52.7	272	2 Q9AMP6	Q9AMP6 ehrlichia s
42	58	52.7	272	2 Q93B54	Q93B54 cowdria rum
43	58	52.7	284	2 Q9AF41	Q9AF41 cowdria rum
44	58	52.7	284	2 Q46327	Q46327 cowdria rum
45	58	52.7	286	2 Q52105	Q52105 ehrlichia c

#### ALIGNMENTS

#### RESULT 1

Q9F472 ID Q9F472 PRELIMINARY; PRT; 278 AA.  
AC Q9F472;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE P28-7.  
GN P28-7.  
OS Ehrlichia canis.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=944;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Jake;  
RX MEDLINE=99242757; PubMed=10225842;  
RA McBride J.W., Yu, Xj, Walker D.H.;  
RT "Molecular cloning of the gene for a conserved major immunoreactive  
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic  
RT antigen.";  
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Jake;  
RX MEDLINE=99242757; PubMed=10974556;  
RA McBride J.W., Yu X.J., Walker D.H.;  
RT "A conserved, transcriptionally active p28 multigene locus of  
RT Ehrlichia canis.";  
RL Gene 254:245-252(2000)  
DR EMBL; AF082744; AAC64550.2; -;  
DR InterPro; IPR002566; Surface\_Ag\_msp4.  
DR Pfam; PF01617; Surface Ag 2; 1.  
SQ SEQUENCE 278 AA; 30455 MW; 2411CAAB4C56CA74 CRC64;

Query Match 100.0%; Score 110; DB 2; Length 278;  
Best Local Similarity 100.0%; Pred. No. 2.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSTGVFGLKHDWDGSPILK 20  
|||||  
Db 60 KSTGVFGLKHDWDGSPILK 79  
|||||

RESULT 2  
Q9R8A8 PRELIMINARY; PRT; 278 AA.  
AC Q9R8A8 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE 28 kDa outer membrane protein (Fragment).  
OS Ehrlichia canis.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=944;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=Oklahoma;  
RC MEDLINE=99242757; PubMed=10225842;  
RX McBride J.W., Yu, Xj, Walker D.H.;  
RA "Molecular cloning of the gene for a conserved major immunoreactive  
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic  
RT antigen."  
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).  
DR EMBL; AF082746; AAC64552.1; -;  
DR InterPro; IPR002566; Surface\_Ag\_msp4.  
DR Pfam; PF01617; Surface\_Ag\_2; 1.  
FT NON TER 278 278  
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 100.0%; Score 110; DB 2; Length 278;  
Best Local Similarity 100.0%; Pred. No. 2.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSTGVFGLKHDWDGSPILK 20  
|||||  
Db 60 KSTGVFGLKHDWDGSPILK 79  
|||||

RESULT 3  
Q9R8A7 PRELIMINARY; PRT; 278 AA.  
AC Q9R8A7 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE 28 kDa outer membrane protein (Fragment).  
OS Ehrlichia canis.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=944;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=Demon;  
RC MEDLINE=99242757; PubMed=10225842;  
RX McBride J.W., Yu, Xj, Walker D.H.;  
RA "Molecular cloning of the gene for a conserved major immunoreactive  
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic  
RT antigen."  
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).  
DR EMBL; AF082747; AAC64553.1; -;  
DR InterPro; IPR002566; Surface\_Ag\_msp4.  
DR Pfam; PF01617; Surface\_Ag\_2; 1.  
FT NON TER 278 278  
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 100.0%; Score 110; DB 2; Length 278;  
Best Local Similarity 100.0%; Pred. No. 2.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSTGVFGLKHDWDGSPILK 20  
|||||

Db 60 KSTGVFGLKHDWDGSPILK 79  
|||||

RESULT 4  
Q9R3J3 PRELIMINARY; PRT; 278 AA.  
AC Q9R3J3 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE 28 kDa outer membrane protein (Fragment).  
OS Ehrlichia canis.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=944;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=Florida;  
RC MEDLINE=99242757; PubMed=10225842;  
RX McBride J.W., Yu, Xj, Walker D.H.;  
RA "Molecular cloning of the gene for a conserved major immunoreactive  
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic  
RT antigen."  
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).  
DR EMBL; AF082750; AAC64556.1; -;  
DR InterPro; IPR002566; Surface\_Ag\_msp4.  
DR Pfam; PF01617; Surface\_Ag\_2; 1.  
FT NON TER 278 278  
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 100.0%; Score 110; DB 2; Length 278;  
Best Local Similarity 100.0%; Pred. No. 2.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSTGVFGLKHDWDGSPILK 20  
|||||  
Db 60 KSTGVFGLKHDWDGSPILK 79  
|||||

RESULT 5  
Q9R8A6 PRELIMINARY; PRT; 278 AA.  
AC Q9R8A6 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE 28 kDa outer membrane protein (Fragment).  
OS Ehrlichia canis.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=944;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=DJ;  
RC MEDLINE=99242757; PubMed=10225842;  
RX McBride J.W., Yu, Xj, Walker D.H.;  
RA "Molecular cloning of the gene for a conserved major immunoreactive  
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic  
RT antigen."  
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).  
DR EMBL; AF082748; AAC64554.1; -;  
DR InterPro; IPR002566; Surface\_Ag\_msp4.  
DR Pfam; PF01617; Surface\_Ag\_2; 1.  
FT NON TER 278 278  
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 100.0%; Score 110; DB 2; Length 278;  
Best Local Similarity 100.0%; Pred. No. 2.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSTGVFGLKHDWDGSPILK 20  
|||||

Db 60 KSTVGVEGLKHDWDGSPILK 79

## RESULT 6

Q9R8A9 PRELIMINARY; PRT; 278 AA.  
AC Q9R8A9  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE 28 kDa outer membrane protein (Fragment).  
OS Ehrlichia canis.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=944;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Louisiana;  
RX MEDLINE=99242757; PubMed=1025842;  
RA McBride J.W., Yu, Xj, Walker D.H.;  
RT "Molecular cloning of the gene for a conserved major immunoreactive  
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic  
RT antigen";  
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).  
DR EMBL: AF082745; AAC64551.1; -;  
DR InterPro: IPR002566; Surface\_Ag\_msp4.  
DR Pfam: PF01617; Surface\_Ag\_2; 1.  
FT NON TER 278 278  
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 100.0%; Score 110; DB 2; Length 278;  
Best Local Similarity 100.0%; Pred. No. 2.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTVGVEGLKHDWDGSPILK 20

Db 60 KSTVGVEGLKHDWDGSPILK 79

## RESULT 7

Q9R8A5 PRELIMINARY; PRT; 278 AA.  
AC Q9R8A5  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE 28 kDa outer membrane protein (Fragment).  
OS Ehrlichia canis.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=944;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Fuzzy;  
RX MEDLINE=99242757; PubMed=1025842;  
RA McBride J.W., Yu, Xj, Walker D.H.;  
RT "Molecular cloning of the gene for a conserved major immunoreactive  
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic  
RT antigen";  
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).  
DR EMBL: AF082749; AAC64551.1; -;  
DR InterPro: IPR002566; Surface\_Ag\_msp4.  
DR Pfam: PF01617; Surface\_Ag\_2; 1.  
FT NON TER 278 278  
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 100.0%; Score 110; DB 2; Length 278;  
Best Local Similarity 100.0%; Pred. No. 2.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTVGVEGLKHDWDGSPILK 20

Db 60 KSTVGVEGLKHDWDGSPILK 79

## RESULT 8

Q9ZGJ1 PRELIMINARY; PRT; 307 AA.  
AC Q9ZGJ1  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Major outer membrane protein P30-1.  
GN P30-1.  
OS Ehrlichia canis.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=944;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Oklahoma;  
RX MEDLINE=98371112; PubMed=9705412;  
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;  
RT "Cloning and characterization of multigenes encoding the  
RT immunodominant 30-kilodalton major outer membrane proteins of  
RT Ehrlichia canis and application of the recombinant protein for  
RT serodiagnosis";  
RL J. Clin. Microbiol. 36:2671-2680(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Oklahoma;  
RX MEDLINE=21153566; PubMed=11254561;  
RA Ohashi N., Rikihisa Y., Unver A.;  
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer  
RT Membrane Protein Multigene Family in Ehrlichia canis and E.  
RT chaffeensis";  
RL Infect. Immun. 69:2083-2091(2001).  
DR EMBL: AF078553; AAC68666.1; -;  
DR InterPro: IPR002566; Surface\_Ag\_msp4.  
DR Pfam: PF01617; Surface\_Ag\_2; 1.  
SQ SEQUENCE 307 AA; 34108 MW; 9DA9FDG3EBF8BC97 CRC64;

Query Match 100.0%; Score 110; DB 2; Length 307;  
Best Local Similarity 100.0%; Pred. No. 3.2e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTVGVEGLKHDWDGSPILK 20

Db 89 KSTVGVEGLKHDWDGSPILK 108

## RESULT 9

Q9ADV3 PRELIMINARY; PRT; 280 AA.  
AC Q9ADV3  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
DE Major outer membrane protein P30-2.  
GN P30-2.  
OS Ehrlichia canis.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=944;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Oklahoma;  
RX MEDLINE=98371112; PubMed=9705412;  
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;  
RT "Cloning and characterization of multigenes encoding the  
RT immunodominant 30-kilodalton major outer membrane proteins of  
RT Ehrlichia canis and application of the recombinant protein for  
RT serodiagnosis";  
RL J. Clin. Microbiol. 36:2671-2680(1998).  
RN [2]  
RP SEQUENCE FROM N.A.

```
RC STRAIN=Oklahoma;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF078553; AAK28699.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30803 MW; 27238BE1C7E68A91 CRC64;

Query Match 71.8%; Score 79; DB 2; Length 280;
Best Local Similarity 82.4%; Pred. No. 0.00023;
Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy . 2 STVGVLKHDWDGSP1 18
Db 60 STVGVLKHDWDGSP1 76

RESULT 10
Q9F473 PRELIMINARY; PRT; 280 AA.
ID Q9F473
AC Q9F473;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P28-6.
GN P28-6.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Jake;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Jake;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis.";
RL Gene 254:245-252(2000).
DR EMBL; AF082744; AAG14361.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30762 MW; BE284A4B94FE3123 CRC64;

Query Match 71.8%; Score 79; DB 2; Length 280;
Best Local Similarity 82.4%; Pred. No. 0.00023;
Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 STVGVLKHDWDGSP1 18
Db 60 STVGVLKHDWDGSP1 76

RESULT 11
O52107 PRELIMINARY; PRT; 280 AA.
ID O52107
AC O52107;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Major outer membrane protein OMP-1F (28kDa outer membrane protein gene
```

```
DE OMP-1F.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RT are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Allemen A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
RT the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas, and Osciola;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; U72291; AAC02940.1; -.
DR EMBL; AF479834; AAC12937.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30731 MW; CCAA6C34E2AF393E CRC64;

Query Match 69.1%; Score 76; DB 2; Length 280;
Best Local Similarity 73.7%; Pred. No. 0.00069;
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 STVGVLKHDWDGSP1K 20
Db 61 TTTGVFLKQDWDGSP1K 79

RESULT 12
Q9RH35 PRELIMINARY; PRT; 246 AA.
ID Q9RH35
AC Q9RH35;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2003 (TrEMBLrel. 23, Last annotation update)
DE Outer membrane protein p28 (Fragment).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=St.Vincent;
RX MEDLINE=99175287; PubMed=10074538;
```





Job time : 98 secs

---



XX PS Claim 1; Page 5; 29pp; English.

XX CC The invention relates to a composition of matter comprising a polypeptide

XX CC isolated from Ehrlichia species. The composition can be used for

XX CC detecting the presence of antibodies to Ehrlichia, comprising contacting

XX CC one or more polypeptides with a test sample suspected of comprising

XX CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody

XX CC complexes to form and detecting the complexes, where the detection of

XX CC polypeptide/antibody complexes is an indication that antibodies to

XX CC Ehrlichia are present in the test sample. The composition is useful for

XX CC detecting or quantifying the presence of E. canis or E. chaffeensis

XX CC infection in mammals. The polypeptides can be used to develop monoclonal

XX CC and/or polyclonal antibodies that can be employed in assay systems and in

XX CC the generation of chimeric antibodies for therapeutic use or other

XX CC similar applications. This sequence represents an E. canis peptide

XX CC fragment used in the composition of the invention.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 110; DB 23; Length 20;

Best Local Similarity 100.0%; Pred. No. 2.5e-11;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSTGVFGLKHDWDGSPILK 20

Db 1 KSTGVFGLKHDWDGSPILK 20

RESULT 2

AA71477

ID AA71477 standard; Protein; 278 AA.

XX AC AA71477;

XX DT 12-OCT-2000 (first entry)

XX DE Ehrlichia canis immunoreactive protein Eca28-1.

XX KW Homologous 28-kDa protein gene; Eca28-1; immunoreactive; vaccine;

XX KW p28 gene; polymorphic multiple gene family; immunoprotective antigen;

XX KW antibacterial; canine ehrlichiosis; canine tropical pancytopenia;

XX KW tick-borne rickettsial disease; serodiagnosis.

XX OS Ehrlichia canis.

FH Key Location/Qualifiers

FT Peptide 1..23 /label= Signal\_peptide

FT Protein 24..278 /label= Mature\_Eca28-1\_28-kDa\_protein

FT Cleavage-site 21..23 /label= Signal\_peptidase\_cleavage\_site

FT Cleavage-site 23..25 /note= "Additional putative signal peptidase cleavage site"

FT Region 27..34 /label= Variable\_region\_1

FT Region 35..76 /note= "contains immunoreactive peptides"

FT Region 76..84 /label= Variable\_region\_2

FT Region 85..148 /note= "contains immunoreactive peptides"

FT Region 148..156 /label= Variable\_region\_3

FT Region 157..246 /note= "contains immunoreactive peptides"

FT Region 246..258 /label= Variable\_region\_4

FT Region 259..278 /note= "contains immunoreactive peptides"

XX PN WO200032745-A2.

XX PD 08-JUN-2000.

PF 24-NOV-1999; 99WO-US28075.

XX 30-NOV-1998; 98US-0201458.

PR 03-MAR-1999; 99US-0261358.

XX (RERE-) RES DEV FOUND.

XX Walker DH, Yu X, McBride JW;

PI WPI; 2000-412298/35.

XX N-PSDB; AAD01292.

DR Ehrlichia canis antigens useful for vaccinating against canine

XX ehrlichiosis in dogs -

PT Claim 12; Fig 1; 86pp; English.

XX The patent relates to homologous 28-kilodalton (kDa) protein genes of

XX Ehrlichia canis, designated Eca28SA1, Eca28SA2, Eca28SA3, Eca28-1 and

XX Eca28-2. These genes are members of a polymorphic multiple gene family

XX and contained in a single locus of 5.92 kb. The 28-kDa proteins are

XX immunoreactive with anti-E. canis serum hence are important

XX CC immunoprotective antigens. The protein is useful for vaccinating

XX CC against E. canis infections such as canine ehrlichiosis in dogs.

XX CC Canine ehrlichiosis, also known as canine tropical pancytopenia, is a

XX CC tick-borne rickettsial disease of dogs. Eca28-1 is conserved amongst

XX CC different strains of E. canis and hence useful for serodiagnosis of

XX CC canine ehrlichiosis. The present sequence is a E. canis

XX CC Eca28-1 30-kDa protein which is post-translationally modified to a

XX CC mature 28-kDa protein by cleavage of N-terminal signal sequence.

XX SQ Sequence 278 AA;

Query Match 100.0%; Score 110; DB 21; Length 278;

Best Local Similarity 100.0%; Pred. No. 5.3e-10;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSTGVFGLKHDWDGSPILK 20

Db 60 KSTGVFGLKHDWDGSPILK 79

RESULT 3

AAU96100

ID AAU96100 standard; Protein; 278 AA.

XX AC AAU96100;

XX DT 02-JUL-2002 (first entry)

XX DE Ehrlichia canis p28-7.

XX KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;

XX KW antibacterial.

XX OS Ehrlichia canis.

XX WO200222782-A2.

XX 21-MAR-2002.

XX 12-SEP-2001; 2001WO-US28759.

XX 12-SEP-2000; 2000US-0660587.

XX (RERE-) RES DEV FOUND.

XX Walker DH, Yu X, McBride JW;

PI WPI; 2002-351882/38.

XX N-PSDB; ABK68852.

XX New recombinant homologous 28 kilodalton immunodominant protein from

```

PT Ehrlichia canis, useful for treating Ehrlichia canis infections
PS Claim 16; Figure 1; 106pp; English.
XX
CC The invention relates to a recombinant homologous 28 kDa immunodominant
CC protein, p28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
CC dispersed in a pharmaceutically acceptable carrier, is useful for
CC inhibiting E. canis infection in a subject. (I) is useful in the
CC development of vaccines and serodiagnostics that are particularly
CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
CC represent the 28-kDa antigen amino acid sequences of the invention.
XX
SQ Sequence 278 AA;
Query Match 100.0%; Score 110; DB 23; Length 278;
Best Local Similarity 100.0%; Pred. No. 5.9e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSTGVFGLKHDWDGSPILK 20
Db 60 KSTGVFGLKHDWDGSPILK 79
RESULT 4
AAU06961
ID AAY06961 standard; Protein; 307 AA.
XX AC AAY06961;
XX
XX 05-JUL-1999 (first entry)
XX
XX E. canis P30-1 protein.
XX
XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
XX detection; dog.
XX
XX Ehrlichia canis.
XX
XX WO9913720-A1.
XX
XX 25-MAR-1999.
XX
XX 18-SEP-1998; 98WO-US19600.
XX
XX 19-SEP-1997; 97US-0059353.
XX (OHIS ) UNIV OHIO STATE.
XX
XX Ohashi N, Rikihisa Y;
XX
XX WPI; 1999-254290/21.
XX N-PSDB; AAX34761.
XX
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
XX Ehrlichia canis
XX
XX Disclosure; Fig 21B; 55pp; English.
XX
XX The invention provides isolated outer membrane proteins (OMP) from
XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
XX in AAY06943-958. The E. canis proteins form part of the P30 family and
XX consist of proteins shown in AAY06959-970. The proteins and genes are
XX used to detect E. chaffeensis in patients and E. canis in dogs.
XX
XX Sequence 307 AA;
Query Match 100.0%; Score 110; DB 20; Length 307;
Best Local Similarity 100.0%; Pred. No. 5.9e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSTGVFGLKHDWDGSPILK 20
Db 1 KSTGVFGLKHDWDGSPILK 108

```

```

Db 89 KSTGVFGLKHDWDGSPILK 108
RESULT 5
ABG77952
ID ABG77952 standard; Protein; 307 AA.
XX AC ABG77952;
XX
XX 15-NOV-2002 (first entry)
XX
XX Ehrlichia canis outer membrane protein (P30F) #3.
XX
XX Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
XX Ehrlichia canis.
XX US2002120115-A1.
XX
XX 29-AUG-2002.
XX
XX 28-JAN-2002; 2002US-0059964.
XX
XX 19-MAY-1999; 99US-0314701.
XX
XX (RIKI/) RIKIHISA Y.
XX (OHAS/) OHASHI N.
XX
XX Rikihisa Y, Ohashi N;
XX
XX WPI; 2002-618954/66.
XX N-PSDB; ABS63293.
XX
XX Isolated polynucleotide encoding an outer membrane protein of E. canis
XX or E. chaffeensis used in the diagnosis of infection -
XX Claim 10; Fig 21B; 49pp; English.
XX
XX The invention relates to an isolated polynucleotide encoding an outer
XX membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
XX in the diagnosis of infection. An infection such as human ehrlichiosis or
XX canine ehrlichiosis can be diagnosed by providing a serum sample from the
XX patient, providing a polypeptide or mixture of polypeptides, contacting
XX the sample with the polypeptide and assaying for the formation of a
XX complex between antibodies in the serum sample and the polypeptide, where
XX formation of a complex is indicative of infection with E. chaffeensis.
XX This sequence represents an Ehrlichia outer membrane protein of the
XX invention.
XX
XX Sequence 307 AA;
Query Match 100.0%; Score 110; DB 23; Length 307;
Best Local Similarity 100.0%; Pred. No. 5.9e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSTGVFGLKHDWDGSPILK 20
Db 89 KSTGVFGLKHDWDGSPILK 108
RESULT 6
AAW51094
ID AAW51094 standard; Protein; 280 AA.
XX AC AAW51094;
XX
XX 14-SEP-1998 (first entry)
XX
XX Ehrlichia chaffeensis VSA4 protein.
XX
XX MAP1 homologue; variable surface antigen; VSA4; rickettsia;
XX DNA vaccine.
XX

```

OS Ehrlichia chaffeensis.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..25  
 FT /note= "putative signal peptide"  
 XX  
 XX WO9816554-A1.  
 XX  
 XX 23-APR-1998.  
 XX  
 XX 17-OCT-1997; 97WO-US19044.  
 XX  
 XX 17-OCT-1996; 96US-0733230.  
 XX  
 XX (UYFL ) UNIV FLORIDA.  
 XX  
 XX Barbet AF, Burr ridge MJ, Ganta RR, Mahan SM, McGuire TC;  
 PI Nyika A, Rurangirwa FR;  
 XX  
 XX WPI; 1998-251232/22.  
 XX  
 XX N-PSDB; AAV07179.  
 XX  
 XX Composition containing nucleic acid encoding rickettsial antigen -  
 PT useful for, e.g. stimulating protective immune response in humans or  
 PT animals  
 PT  
 XX  
 XX Claim 3; Fig 2B; 39pp; English.  
 XX  
 CC This is the full-length variable surface antigen VSA4 protein of  
 CC Ehrlichia chaffeensis. Its amino acid sequence was deduced from a  
 CC partial open reading frame (ORF4) of a genomic locus (see AAV07179)  
 CC of E. chaffeensis that was obtained on the basis of homology to the  
 CC major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.  
 CC This genomic locus included 5 ORFs encoding similar, but  
 CC non-identical proteins (see AAW51091-95). A claimed composition  
 CC comprises a nucleic acid (see AAV07176-82) encoding a polypeptide  
 CC (see AAW51088-99) that elicits a protective immune response against a  
 CC rickettsial pathogen. The nucleic acid is used, in human or  
 CC veterinary medicine, in vaccines to protect against Rickettsia.  
 CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic  
 CC polypeptides can be used diagnostically to detect antibodies  
 CC associated with Ehrlichia infection (claimed).  
 XX  
 XX Sequence 280 AA;  
 SQ  
 Query Match 71.8%; Score 79; DB 19; Length 280;  
 Best Local Similarity 73.7%; Pred. No. 7.8e-05;  
 Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 STGVGFLKHDWGSPTLK 20  
 DB 61 TTIGVFLKQDWGSTISK 79  
 RESULT 7  
 AAY06962  
 ID AAY06962 standard; Protein; 280 AA.  
 XX  
 AC AAY06962;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE E. canis P30-2 protein.  
 XX  
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 KW detection; dog.  
 XX  
 OS Ehrlichia canis.  
 XX  
 XX WO9913720-A1.  
 XX  
 XX 25-MAR-1999.  
 PD  
 XX

PF 18-SEP-1998; 98WO-US19600.  
 XX  
 PR 19-SEP-1997; 97US-0059353.  
 XX  
 PA (OHIS ) UNIV OHIO STATE.  
 XX  
 XX Ohashi N, Rikihisa Y;  
 XX  
 XX WPI; 1999-254290/21.  
 DR  
 DR N-PSDB; AAX34762.  
 XX  
 XX Novel outer membrane proteins from Ehrlichia chaffeensis and  
 PT Ehrlichia canis  
 PT  
 XX  
 PS Disclosure; Fig 22B; 55pp; English.  
 XX  
 CC The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1B to 2) shown  
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and  
 CC consist of proteins shown in AAY06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX  
 XX Sequence 280 AA;  
 SQ  
 Query Match 71.8%; Score 79; DB 20; Length 280;  
 Best Local Similarity 82.4%; Pred. No. 7.8e-05;  
 Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 STGVGFLKHDWGSPI 18  
 DB 60 STGVGFLKHDWNGGTI 76  
 RESULT 8  
 AAB36188  
 ID AAB36188 standard; Protein; 280 AA.  
 XX  
 AC AAB36188;  
 XX  
 DT 02-MAR-2001 (first entry)  
 XX  
 DE Ehrlichia chaffeensis partial VSA4.  
 XX  
 KW Ehrlichia chaffeensis; VSA4; variable surface antigen 4; MAP1;  
 KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;  
 KW Rickettsia; Cowdria; Anaplasma; map2; inhorrif3; 4horrif1; 18horrif1;  
 KW 3gdrorf3.  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 XX WO200065063-A2.  
 XX  
 PD 02-NOV-2000.  
 XX  
 PF 21-APR-2000; 2000WO-US10886.  
 XX  
 PR 22-APR-1999; 99US-0130725.  
 XX  
 XX (UYFL ) UNIV FLORIDA.  
 XX  
 XX Barbet AF, Bowie MV, Ganta RR, Burr ridge MJ, Mahan SM, McGuire TC;  
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;  
 XX  
 XX WPI; 2000-679675/66.  
 DR  
 DR N-PSDB; AAC68705.  
 XX  
 XX New polynucleotides useful as DNA vaccines for conferring immunity to  
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,  
 PT comprises major antigenic protein 1.or'2 gene of rickettsial pathogens  
 PT  
 XX  
 PS Claim 3; Page 45-46; 63pp; English.

XX The present sequence shows a high degree of similarity to the major  
 CC antigenic protein 1 (MAP1) of *Ehrlichia* sp. The MAP1 gene may be  
 CC used in a vaccines to protect animals or humans against rickettsial  
 CC diseases caused by a organisms of *Rickettsia* sp., *Ehrlichia* sp.,  
 CC *Anaplasma* sp. or *Cowdria* sp. The vaccine elicits an immune response  
 CC protective against the rickettsial pathogen. The nucleic acid vaccines  
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.  
 CC *Cowdria* ruminatum genes designated map 2, ihworf3, 4hworf1, 1shworf1  
 CC and 3gdotf3 may be used in therapeutic and diagnostic applications. The  
 CC polypeptides are useful for detecting antibodies associated with  
 CC infection by a rickettsial pathogen whilst the polynucleotides may be  
 CC used to detect the presence of rickettsial nucleic acids.

SQ Sequence 280 AA;  
 Query Match 71.8%; Score 79; DB 21; Length 280;  
 Best Local Similarity 73.7%; Pred. No. 7.8e-05;  
 Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 STVGVFGLKHDWDGSPILK 20  
 :||| ||||| ||||| |||||  
 Db 61 TTIGVFGLKQDWDGSTISK 79

RESULT 9  
 AAY71479  
 ID AAY71479 standard; Protein; 280 AA.

AC AAY71479;

DT 12-OCT-2000 (first entry)

DE *Ehrlichia canis* immunoreactive protein Eca28SA3.

XX Homologous 28-kDa protein gene; Eca28SA3; immunoreactive; vaccine;  
 KW p28 gene; polymorphic multiple gene family; immunoprotective antigen;  
 KW antibacterial; canine ehrlichiosis; canine tropical pancytopenia;  
 KW tick-borne rickettsial disease; serodiagnosis.

XX *Ehrlichia canis*.

Key Location/Qualifiers

FT Peptide 1..23

FT /label= Signal\_peptide

FT Protein 24..280

FT /label= Mature\_Eca28SA3\_28-kDa\_protein

XX W0200032745-A2.

XX 08-JUN-2000.

XX 24-NOV-1999; 99WO-US28075.

XX 30-NOV-1998; 98US-0201458.

PR 03-MAR-1999; 99US-0261358.

XX (SERE-) RES DEV FOUND.

XX Walker DH, Yu X, McBride JW;

XX WPI; 2000-412298/35.

DR N-PSDB; AAD01294, AAD01295.

PT *Ehrlichia canis* antigens useful for vaccinating against canine  
 PT ehrlichiosis in dogs -

PS Claim 12; Page 68-69; 86pp; English.

XX The patent relates to homologous 28-kiloDalton (kDa) protein genes of  
 CC *Ehrlichia canis*, designated Eca28SA1, Eca28SA2, Eca28SA3, Eca28-1 and  
 CC Eca28-2. These genes are members of a polymorphic multiple gene family  
 CC and contained in a single locus of 5.592 kb. The 28-kDa proteins are

CC immunoreactive with anti-*E. canis* serum hence are important  
 CC immunoprotective antigens. The protein is useful for vaccinating  
 CC against *E. canis* infections such as canine ehrlichiosis in dogs.  
 CC Canine ehrlichiosis, also known as canine tropical pancytopenia, is a  
 CC tick-borne rickettsial disease of dogs. Eca28-1 is conserved amongst  
 CC different strains of *E. canis* and hence useful for serodiagnosis of  
 CC canine ehrlichiosis. The present sequence is a *E. canis*  
 CC Eca28SA3 30-kDa protein which is post-translationally modified to a  
 CC mature 28-kDa protein by cleavage of N-terminal signal sequence.

SQ Sequence 280 AA;

Query Match 71.8%; Score 79; DB 21; Length 280;  
 Best Local Similarity 82.4%; Pred. No. 7.8e-05;  
 Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 STVGVFGLKHDWDGSP1 18

||||| ||||| ||||| |||||

Db 60 STVGVFGLKHDWNGGTI 76

RESULT 10

AAU04198

ID AAU04198 standard; Protein; 280 AA.

AC AAU04198;

DT 23-OCT-2001 (first entry)

DE Variable surface antigen 4 (VSA4) from *Ehrlichia chaffeensis*.

XX Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;  
 KW infection; heartwater; diagnostic; variable surface antigen; VSA.

XX *Ehrlichia chaffeensis*.

XX US6251872-B1.

XX 26-JUN-2001.

XX 17-OCT-1997; 97US-0953326.

XX 17-OCT-1996; 96US-0733230.

XX (UYFL) UNIV FLORIDA.

PI Barbet AF, Ganta RR, McGuire TC, Burrige MJ, Nyika A;

PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;

XX WPI; 2001-424487/45.

XX N-PSDB; AAS07578.

XX New MAP2 genes and polypeptides useful as vaccines for conferring  
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or  
 PT as molecular markers in nucleic acid analysis procedures -

XX Example 3; Fig 2A-2B; 30pp; English.

XX The sequence represents the amino acid sequence of variable surface  
 CC antigen 4 (VSA4) isolated from *Ehrlichia chaffeensis*, which  
 CC has similarity to major antigen protein (MAP). The MAP polynucleotides  
 CC and polypeptides are useful as vaccines for conferring immunity to  
 CC rickettsia infection, including *Cowdria ruminatum* causing heartwater.  
 CC The MAP polynucleotides may be used as molecular markers in nucleic acid  
 CC analysis procedures, and to produce the MAP polypeptides, which may  
 CC be used to raise antibodies that are reactive with the polypeptides.  
 CC The nucleic acids may further be used as probes to identify  
 CC complementary sequences within other nucleic acid molecules or genomes,  
 CC where such probes can be applied to identify or distinguish infectious  
 CC strains of organisms in diagnostic procedures or in rickettsial  
 CC research where identification of particular organisms or strains is  
 CC needed.

XX

SQ Sequence 280 AA;

Query Match 71.8%; Score 79; DB 22; Length 280;  
 Best Local Similarity 73.7%; Pred. No. 7.8e-05;  
 Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 STVGVFGLKHDWDGSPILK 20  
 :|||:|||||:|||||:|  
 Db 61 TTIGVFGLKQDWDGSTISK 79

RESULT 11  
 ABG77953  
 ID ABG77953 standard; Protein; 280 AA.  
 AC ABG77953;  
 XX  
 DT 15-NOV-2002 (first entry)  
 XX  
 DE Ehrlichia canis outer membrane protein (P30F) #4.  
 XX  
 KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.  
 XX  
 OS Ehrlichia canis.  
 XX  
 PN US2002120115-A1.  
 XX  
 PD 29-AUG-2002.  
 XX  
 PF 28-JAN-2002; 2002US-0059964.  
 XX  
 PR 19-MAY-1999; 99US-0314701.  
 XX  
 PA (RIKI/) RIKIHISA Y.  
 PA (OHAS/) OHASHI N.  
 XX  
 PI Rikihisa Y, Ohashi N;  
 XX  
 DR WPI; 2002-618954/66.  
 DR N-PSDB; ABS63294.  
 XX  
 PT Isolated polynucleotide encoding an outer membrane protein of E.canis  
 or E.chaffeensis used in the diagnosis of infection -  
 XX  
 PS Claim 10; Fig 22B; 49pp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide encoding an outer  
 membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used  
 in the diagnosis of infection. An infection such as human ehrlichiosis or  
 canine ehrlichiosis can be diagnosed by providing a serum sample from the  
 patient, providing a polypeptide or mixture of polypeptides, contacting  
 the sample with the polypeptide and assaying for the formation of a  
 complex between antibodies in the serum sample and the polypeptide, where  
 formation of a complex is indicative of infection with E. chaffeensis.  
 CC This sequence represents an Ehrlichia outer membrane protein of the  
 invention.  
 XX  
 SQ Sequence 280 AA;

Query Match 71.8%; Score 79; DB 23; Length 280;  
 Best Local Similarity 82.4%; Pred. No. 7.8e-05;  
 Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 STVGVFGLKHDWDGSPIL 18  
 :|||||:|||||:|  
 Db 60 STVGVFGLKHDWNGGTI 76

RESULT 12  
 AAU96102  
 ID AAU96102 standard; Protein; 280 AA.  
 XX  
 AC AAU96102;

XX  
 DT 02-JUL-2002 (first entry)  
 XX  
 DE Ehrlichia canis p28-6.  
 XX  
 KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;  
 KW antibacterial.  
 XX  
 OS Ehrlichia canis.  
 XX  
 PN WO200222782-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PF 12-SEP-2001; 2001WO-US28759.  
 XX  
 PR 12-SEP-2000; 2000US-0660587.  
 XX  
 PA (RERE-) RES DEV FOUND.  
 XX  
 PI Walker DH, Yu X, McBride JW;  
 XX  
 DR WPI; 2002-351882/38.  
 DR N-PSDB; ABK68854.  
 XX  
 PT New recombinant homologous 28 kilodalton immunodominant protein from  
 Ehrlichia canis, useful for treating Ehrlichia canis infections -  
 XX  
 PS Claim 16; Figure 7; 106pp; English.  
 XX  
 CC The invention relates to a recombinant homologous 28 kDa immunodominant  
 protein, P28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably  
 dispersed in a pharmaceutically acceptable carrier, is useful for  
 inhibiting E. canis infection in a subject. (I) is useful in the  
 development of vaccines and serodiagnostics that are particularly  
 effective for disease prevention and serodiagnosis. AAU96100-AAU96118  
 CC represent the 28-kDa antigen amino acid sequences of the invention.  
 XX  
 SQ Sequence 280 AA;

Query Match 71.8%; Score 79; DB 23; Length 280;  
 Best Local Similarity 82.4%; Pred. No. 7.8e-05;  
 Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 STVGVFGLKHDWDGSPIL 18  
 :|||||:|||||:|  
 Db 60 STVGVFGLKHDWNGGTI 76

RESULT 13  
 AAU73417  
 ID AAU73417 standard; Protein; 280 AA.  
 XX  
 AC AAU73417;  
 XX  
 DT 12-MAR-2002 (first entry)  
 XX  
 DE Ehrlichia chaffeensis outer membrane protein P28-18.  
 XX  
 KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 PN WO200183699-A2.  
 XX  
 PD 08-NOV-2001.  
 XX  
 PF 01-MAY-2001; 2001WO-US13997.  
 XX  
 PR 01-MAY-2000; 2000US-201035P.  
 XX  
 PA (RERE-) RES DEV FOUND.  
 XX



PI Walker DH, Yu X;  
XX WPI; 2002-066527/09.  
XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated  
PT P28 useful as a vaccine against Ehrlichia chaffeensis  
XX  
XX Disclosure; Figure 2; 97pp; English.  
XX  
XX The invention relates to isolated and purified 28-kDa outer membrane  
CC proteins (p28-1 to p28-21) of Ehrlichia chaffeensis. P28 proteins  
CC are encoded by a 28kDa outer membrane protein multigene family. P28  
CC proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28  
CC is useful for transfecting a host cell. AAU73400-AAU73420 represent  
CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.  
XX  
XX Sequence 280 AA;  
SQ  
  
Query Match 71.8%; Score 79; DB 23; Length 280;  
Best Local Similarity 73.7%; Pred. No. 7.8e-05;  
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 2 STGVFGLKHDWDGSPILK 20  
Db 61 TTIGVFGLKQDWDGSTISK 79  
  
RESULT 14  
AAU06948  
ID AAY06948 standard; Protein; 280 AA.  
XX  
AC AAY06948;  
XX  
XX 05-JUL-1999 (first entry)  
XX  
XX E. chaffeensis OMP-1F protein.  
XX  
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
XX detection; dog.  
XX  
OS Ehrlichia chaffeensis.  
XX  
XX WO9913720-A1.  
XX  
XX 25-MAR-1999.  
XX  
XX 18-SEP-1998; 98WO-US19600.  
XX  
XX 19-SEP-1997; 97US-0059353.  
XX  
XX (CHIS ) UNIV OHIO STATE.  
XX  
XX Ohashi N, Rikihisa Y;  
XX  
XX WPI; 1999-254290/21.  
DR N-PSDB; AAX34748.  
XX  
XX Novel outer membrane proteins from Ehrlichia chaffeensis and  
PT Ehrlichia canis  
XX  
XX Claim 16; Fig 8B; 55pp; English.  
XX  
XX The invention provides isolated outer membrane proteins (OMP) from  
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
CC in AAU06943-958. The E. canis proteins form part of the P30 family and  
CC consist of proteins shown in AAY06959-970. The proteins and genes are  
CC used to detect E. chaffeensis in patients and E. canis in dogs.  
XX  
XX Sequence 280 AA;  
SQ  
  
Query Match 69.1%; Score 76; DB 20; Length 280;  
Best Local Similarity 73.7%; Pred. No. 0.00025;  
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 2 STGVFGLKHDWDGSPILK 20  
Db 61 TTIGVFGLKQDWDGSTISK 79  
  
RESULT 15  
ABG77940  
ID ABG77940 standard; Protein; 280 AA.  
XX  
XX ABG77940;  
XX  
XX 15-NOV-2002 (first entry)  
DT  
XX Ehrlichia chaffeensis outer membrane protein (OMP) #6.  
XX  
XX Outer membrane protein; OMP; P30F; ehrlichiosis; infection.  
XX  
XX Ehrlichia chaffeensis.  
XX  
XX US2002120115-A1.  
XX  
XX 29-AUG-2002.  
XX  
XX 28-JAN-2002; 2002US-0059964.  
XX  
XX 19-MAY-1999; 99US-0314701.  
XX  
XX (RIKI/) RIKIHISA Y.  
XX (OHAS/) OHASHI N.  
XX  
XX Rikihisa Y, Ohashi N;  
XX  
XX WPI; 2002-618954/66.  
DR N-PSDB; ABS83281.  
XX  
XX Isolated polynucleotide encoding an outer membrane protein of E.canis  
PT or E.chaffeensis used in the diagnosis of infection -  
XX  
XX Disclosure; Fig 8B; 49pp; English.  
XX  
XX The invention relates to an isolated polynucleotide encoding an outer  
CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used  
CC in the diagnosis of infection. An infection such as human ehrlichiosis or  
CC canine ehrlichiosis can be diagnosed by providing a serum sample from the  
CC patient, providing a polypeptide or mixture of polypeptides, contacting  
CC the sample with the polypeptide and assaying for the formation of a  
CC complex between antibodies in the serum sample and the polypeptide, where  
CC formation of a complex is indicative of infection with E. chaffeensis.  
CC This sequence represents an Ehrlichia outer membrane protein of the  
CC invention.  
XX  
XX Sequence 280 AA;  
SQ  
  
Query Match 69.1%; Score 76; DB 23; Length 280;  
Best Local Similarity 73.7%; Pred. No. 0.00025;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 2 STGVFGLKHDWDGSPILK 20  
Db 61 TTIGVFGLKQDWDGSTISK 79  
  
Search completed: September 5, 2003, 14:06:37  
Job time : 83 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2003, 14:05:34 ; Search time 28 Seconds  
(without alignments)  
30.222 Million cell updates/sec

Title: US-10-054-354-1

Perfect score: 110

Sequence: 1 KSTVGVFGLKHDWDGSPILK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	278	4	US-09-660-587-2
2	110	100.0	278	4	US-09-261-358A-2
3	110	100.0	278	4	US-09-201-458-2
4	110	100.0	307	4	US-09-314-701-36
5	79	71.8	280	3	US-08-953-326-17
6	79	71.8	280	4	US-09-660-587-6
7	79	71.8	280	4	US-09-261-358A-6
8	79	71.8	280	4	US-09-314-701-38
9	76	69.1	280	4	US-09-660-587-14
10	76	69.1	280	4	US-09-261-358A-14
11	76	69.1	280	4	US-09-201-458-10
12	76	69.1	280	4	US-09-314-701-12
13	72	65.5	276	3	US-08-953-326-18
14	72	65.5	280	3	US-08-733-230-4
15	72	65.5	280	3	US-08-953-326-4
16	72	65.5	281	4	US-09-660-587-9
17	72	65.5	281	4	US-09-261-358A-9
18	72	65.5	281	4	US-09-201-458-5
19	72	65.5	281	4	US-09-314-701-2
20	70	63.6	288	4	US-09-314-701-32
21	60	54.5	287	3	US-08-733-230-2
22	60	54.5	287	3	US-08-953-326-2
23	58	52.7	284	4	US-09-660-587-15
24	58	52.7	284	4	US-09-261-358A-15
25	58	52.7	284	4	US-09-201-458-11
26	58	52.7	286	3	US-08-953-326-15
27	58	52.7	286	3	US-09-660-587-12

28	58	52.7	286	4	US-09-261-358A-12	Sequence 12, Appl
29	58	52.7	286	4	US-09-201-458-8	Sequence 8, Appl
30	58	52.7	286	4	US-09-314-701-8	Sequence 8, Appl
31	57	51.8	276	4	US-09-660-587-44	Sequence 44, Appl
32	57	51.8	276	4	US-09-314-701-42	Sequence 42, Appl
33	53	48.2	278	3	US-08-953-326-16	Sequence 16, Appl
34	53	48.2	278	4	US-09-660-587-13	Sequence 13, Appl
35	53	48.2	278	4	US-09-261-358A-13	Sequence 13, Appl
36	53	48.2	278	4	US-09-201-458-9	Sequence 9, Appl
37	53	48.2	278	4	US-09-314-701-10	Sequence 10, Appl
38	52	47.3	280	4	US-09-660-587-11	Sequence 11, Appl
39	52	47.3	280	4	US-09-261-358A-11	Sequence 11, Appl
40	52	47.3	280	4	US-09-201-458-7	Sequence 7, Appl
41	52	47.3	280	4	US-09-314-701-6	Sequence 6, Appl
42	48	43.6	133	3	US-08-953-326-20	Sequence 20, Appl
43	48	43.6	133	4	US-09-660-587-7	Sequence 7, Appl
44	48	43.6	133	4	US-09-261-358A-7	Sequence 7, Appl
45	48	43.6	133	4	US-09-201-458-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1

US-09-660-587-2  
; Sequence 2, Application US/0960587  
; Patent No. 6392023  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: US/09/660,587  
; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 2  
; LENGTH: 278  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. canis p28-7 protein  
US-09-660-587-2

Query Match 100.0%; Score 110; DB 4; Length 278;  
Best Local Similarity 100.0%; Pred. No. 78-11;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSTVGVFGLKHDWDGSPILK 20  
Db 60 KSTVGVFGLKHDWDGSPILK 79

RESULT 2

US-09-261-358A-2  
; Sequence 2, Application US/09261358A  
; Patent No. 6403780  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
; FILE REFERENCE: D6152CIP  
; CURRENT APPLICATION NUMBER: US/09/261,358A  
; CURRENT FILING DATE: 1999-03-03  
; PRIOR APPLICATION NUMBER: 09/201,458  
; PRIOR FILING DATE: 1998-11-30  
; NUMBER OF SEQ ID NOS: 33  
; SEQ ID NO 2



RESULT 7  
US-09-261-358A-6  
; Sequence 6, Application US/09261358A  
; Patent No. 6403780  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP  
; CURRENT APPLICATION NUMBER: US/09/261,358A  
; CURRENT FILING DATE: 1999-03-03  
; PRIOR APPLICATION NUMBER: 09/201,458  
; PRIOR FILING DATE: 1998-11-30  
; NUMBER OF SEQ ID NOS: 33  
; SEQ ID NO 6  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of Eca28SA3 protein  
US-09-261-358A-6

Query Match 71.8%; Score 79; DB 4; Length 280;  
Best Local Similarity 82.4%; Pred. No. 1.3e-05;  
Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 STGVFGLKHDWGSPI 18  
| | | | | | | | | | | | | | | | | |  
Db 60 STGVFGLKHDWNGGTI 76

RESULT 8  
US-09-314-701-38  
; Sequence 38, Application US/09314701  
; Patent No. 6544517  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohsei, No. 6544517io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/09/314,701  
; CURRENT FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 38  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
US-09-314-701-38

Query Match 71.8%; Score 79; DB 4; Length 280;  
Best Local Similarity 82.4%; Pred. No. 1.3e-05;  
Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 STGVFGLKHDWGSPI 18  
| | | | | | | | | | | | | | | | | |  
Db 60 STGVFGLKHDWNGGTI 76

RESULT 9  
US-09-660-587-14  
; Sequence 14, Application US/09660587  
; Patent No. 6392023  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein

; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: US/09/660,587  
; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 14  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F  
US-09-660-587-14

Query Match 69.1%; Score 76; DB 4; Length 280;  
Best Local Similarity 73.7%; Pred. No. 4.2e-05;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 STGVFGLKHDWGSPIK 20  
: | | | | | | | | | | | | | | | | | |  
Db 61 TTGTGVLKQDWDGTSISK 79

RESULT 10  
US-09-261-358A-14  
; Sequence 14, Application US/09261358A  
; Patent No. 6403780  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP  
; CURRENT APPLICATION NUMBER: US/09/261,358A  
; CURRENT FILING DATE: 1999-03-03  
; PRIOR APPLICATION NUMBER: 09/201,458  
; PRIOR FILING DATE: 1998-11-30  
; NUMBER OF SEQ ID NOS: 33  
; SEQ ID NO 14  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F  
US-09-261-358A-14

Query Match 69.1%; Score 76; DB 4; Length 280;  
Best Local Similarity 73.7%; Pred. No. 4.2e-05;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 STGVFGLKHDWGSPIK 20  
: | | | | | | | | | | | | | | | | | |  
Db 61 TTGTGVLKQDWDGTSISK 79

RESULT 11  
US-09-201-458-10  
; Sequence 10, Application US/09201458A  
; Patent No. 6458942  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia  
; FILE REFERENCE: D6152  
; CURRENT APPLICATION NUMBER: US/09/201,458A  
; CURRENT FILING DATE: 1998-11-30  
; NUMBER OF SEQ ID NOS: 21  
; SEQ ID NO 10  
; LENGTH: 280

; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F  
US-09-201-458-10

Query Match 69.1%; Score 76; DB 4; Length 280;  
Best Local Similarity 73.7%; Pred. No. 4.2e-05;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 STGVFGLKHDWDGSPILK 20  
:|||||:|||||  
Db 61 TTGVFGLKQWDGSGTISK 79

RESULT 12  
US-09-314-701-12  
; Sequence 12, Application US/09314701  
; Patent No. 6544517  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ganta, Roman R.  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/09/314,701  
; CURRENT FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
US-09-314-701-12

Query Match 69.1%; Score 76; DB 4; Length 280;  
Best Local Similarity 73.7%; Pred. No. 4.2e-05;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 STGVFGLKHDWDGSPILK 20  
:|||||:|||||  
Db 61 TTGVFGLKQWDGSGTISK 79

RESULT 13  
US-08-953-326-18  
; Sequence 18, Application US/08953326  
; Patent No. 6251872  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Ganta, Roman R.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Nyika, Aceme  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Mahan, Suman M.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of  
; TITLE OF INVENTION: Animals and Humans  
; FILE REFERENCE: UF-167C1  
; CURRENT APPLICATION NUMBER: US/08/953,326  
; CURRENT FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/953,326  
; EARLIER FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/733,230  
; EARLIER FILING DATE: 1996-10-17  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 276  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
US-08-953-326-18

Query Match 65.5%; Score 72; DB 3; Length 276;  
Best Local Similarity 76.5%; Pred. No. 0.0002;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 STGVFGLKHDWDGSP1 18  
:|||||:|||||  
Db 60 TTGVFGLKQWDGSAI 76

RESULT 14  
US-08-733-230-4  
; Sequence 4, Application US/08733230  
; Patent No. 6025338  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Ganta, Roman Reddy  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Nyika, Aceme  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Mahan, Suman M.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against  
; TITLE OF INVENTION: Rickettsial Diseases and Methods of Use  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/733,230  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Whitlock, Ted W.  
; REGISTRATION NUMBER: 36,965  
; REFERENCE/DOCKET NUMBER: UF-167  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 352-375-8100  
; TELEFAX: 352-372-5800  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 280 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-733-230-4

Query Match 65.5%; Score 72; DB 3; Length 280;  
Best Local Similarity 76.5%; Pred. No. 0.0002;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 STGVFGLKHDWDGSP1 18  
:|||||:|||||  
Db 61 TTGVFGLKQWDGSAI 77

RESULT 15  
US-08-953-326-4  
; Sequence 4, Application US/08953326  
; Patent No. 6251872  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Ganta, Roman R.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Burridge, Michael J.

; APPLICANT: Nyika, Aceme  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Mahan, Suman M.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of  
; TITLE OF INVENTION: Animals and Humans  
; FILE REFERENCE: UF-167C1  
; CURRENT APPLICATION NUMBER: US/08/953,326  
; CURRENT FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/953,326  
; EARLIER FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/733,230  
; EARLIER FILING DATE: 1996-10-17  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
US-08-953-326-4

Query Match 65.5%; Score 72; DB 3; Length 280;  
Best Local Similarity 76.5%; Pred. No. 0.0002;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 2 STVGVFGLKHDWDSPI 18  
Db 61 TTGVFGLKQNWDSAI 77

Search completed: September 5, 2003, 14:10:12  
Job time : 29 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2003, 14:01:59 ; Search time 40 Seconds  
(without alignments)  
48.084 Million cell updates/sec

Title: US-10-054-354-1  
Perfect score: 110  
Sequence: 1 KSTVGFLKQHDGSPILK 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76.\*  
1: PIR1.\*  
2: PIR2.\*  
3: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79	71.8	280	2 JE0217	28k surface antige
2	72	65.5	276	2 JE0218	28k surface antige
3	58	52.7	284	2 I40882	major antigenic pr
4	58	52.7	286	2 JE0219	28k surface antige
5	53	48.2	278	2 JE0216	28k surface antige
6	50	45.5	756	2 T20109	hypothetical prote
7	48	43.6	133	2 JE0221	28k surface antige
8	47	42.7	396	2 A59236	tauripine dehydrog
9	46	41.8	385	2 A97634	hypothetical prote
10	46	41.8	385	2 AC2857	conserved hypotet
11	45.5	41.4	231	2 T32146	hypothetical prote
12	45	40.9	103	2 F75010	hypothetical prote
13	45	40.9	132	2 B89834	hypothetical prote
14	45	40.9	311	2 E86436	teichoic acid bios
15	45	40.9	570	2 S35137	F28K20.3 protein -
16	45	40.9	570	2 G86777	probable phosphogl
17	45	40.9	704	1 DJBPT7	dihydroxy-acid deh
18	45	40.9	704	2 S07512	DNA-directed DNA p
19	45	40.9	819	2 B87580	conserved hypotet
20	45	40.9	5188	2 B85547	probable RTX famil
21	45	40.9	5291	2 F90696	hypothetical prote
22	44	40.0	130	2 C88102	protein w09G10.6 l
23	44	40.0	160	2 A75466	2-demethylmenaquin
24	44	40.0	301	2 D83445	hypothetical prote
25	44	40.0	340	2 G72596	hypothetical prote
26	44	40.0	348	2 AE3307	hypothetical prote
27	44	40.0	368	2 E64142	hypothetical prote
28	44	40.0	410	2 B59103	hypothetical prote
29	44	40.0	646	2 T48902	sulfate transporte

30	44	40.0	658	2 T48901	sulfate transporte
31	44	40.0	658	2 T49069	sulfate transporte
32	44	40.0	703	2 T51161	hypothetical prote
33	44	40.0	768	2 E86417	unknown protein, 5
34	44	40.0	770	2 T02318	hypothetical prote
35	44	40.0	886	2 C87031	Cell division prot
36	44	40.0	895	2 T45738	hypothetical prote
37	43	39.1	200	2 T50442	hypothetical prote
38	43	39.1	255	2 S66068	conserved hypotet
39	43	39.1	257	2 D89815	conserved hypotet
40	43	39.1	264	2 G81252	NADH2 dehydrogenas
41	43	39.1	265	2 D71838	nadh oxidoreductas
42	43	39.1	286	2 F64677	NADH2 dehydrogenas
43	43	39.1	306	2 T44684	hypothetical prote
44	43	39.1	334	2 F75539	branched-chain ami
45	43	39.1	386	2 AB2455	glycosyltransferas

ALIGNMENTS

RESULT 1

JE0217  
28k surface antigen 4 - Ehrlichia chaffensis  
N/Alternate names: MAP1  
C/Species: Ehrlichia chaffensis  
C/Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 17-Mar-1999  
C/Accession: JE0217  
R/Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrige, M.J.; Alleman, A.R.; Biochem. Biophys. Res. Commun. 247, 636-643, 1998  
A/Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe A/Reference number: JE0216; MUID:98321180; PMID:9647746  
A/Accession: JE0217  
A/Molecule type: DNA  
A/Residues: 1-280 <RED>  
A/Cross-references: GB:AF062761

Query Match 71.8%; Score 79; DB 2; Length 280;  
Best Local Similarity 73.7%; Pred. No. 6.3e-05;  
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 STVGFLKQHDGSPILK 20  
Db 61 TTIGVFLKQDWDGSTIK 79  
:|:|||||

RESULT 2

JE0218  
28k surface antigen 5 - Ehrlichia chaffensis  
N/Alternate names: MAP1  
C/Species: Ehrlichia chaffensis  
C/Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 17-Mar-1999  
C/Accession: JE0218  
R/Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrige, M.J.; Alleman, A.R.; Biochem. Biophys. Res. Commun. 247, 636-643, 1998  
A/Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe A/Reference number: JE0216; MUID:98321180; PMID:9647746  
A/Accession: JE0218  
A/Molecule type: DNA  
A/Residues: 1-276 <RED>  
A/Cross-references: GB:AF062761

Query Match 65.5%; Score 72; DB 2; Length 276;  
Best Local Similarity 76.5%; Pred. No. 0.00081;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 STVGFLKQHDGSPILK 18  
Db 60 TTIGVFLKQDWDGSTIK 76  
:|:|||||

RESULT 3

I40882

major antigenic protein - heartwater rickettsia  
C/Species: Cowdria ruminantium (heartwater rickettsia)  
C/Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 08-Oct-1999  
C/Accession: I40882; S42827  
R/van Vliet, A.H.; Jongejans, F.; van Kleef, M.; van der Zeijst, B.A.  
Infect. Immun. 62, 1451-1456, 1994  
A/Title: Molecular cloning, sequence analysis, and expression of the gene encoding the  
A/Reference number: I40882; MUID:94178956; PMID:8132352  
A/Accession: I40882  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-284 <RES>  
A/Cross-references: EMBL:X74250; NTD:9454266; PIDN:CAA52309.1; PID:9454267  
C/Genetics:  
A/Gene: map1

```
Query Match      52.7%; Score 58; DB 2; Length 284;
Best Local Similarity 66.7%; Pred. No. 0.14;
Matches 10: Conservative 1; Mismatches 4; Indels 0; Gaps 0;
```

Qy 1 KSTVGVFGLKHDWDG 15  
: | | | | | | | |  
Db 61 RDTKAVFGLKKDWDG 75

RESULT 4  
JE0219  
28k surface antigen 2 - Ehrlichia chaffensis  
N/Alternate names: MAP1  
C/Species: Ehrlichia chaffensis  
C/Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 17-Mar-1999  
C/Accession: JE0219  
R/Reedy, G.R.; Sulisona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.  
Biochem. Biophys. Res. Commun. 247, 636-643, 1998  
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe  
A:Reference number: JE0216; MUID:98321180; PMID:9647746  
A:Accession: JE0219  
A:Molecule type: DNA  
A:Residues: 1-286 <RED>  
A:Cross-references: GB:AF062761

```

Query Match      52.7%; Score 58; DB 2; Length 286;
Best Local Similarity 69.2%; Pred. No. 0.14;
Matches 9: Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

Qy	2	STVGVFGLKHDWD	14
		: : : : : :	
Db	61	TTVGVGEGIEQDWD	73

RESULT 5  
JEO216  
28k surface antigen 3 - Ehrlichia chaffensis  
N:Alternate names: MAP1  
C:Species: Ehrlichia chaffensis  
C:Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 17-Mar-1999  
C:Accession: JEO216  
R:R;Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; BurrIDGE, M.J.; Alleman, A.R.  
Biochem. Biophys. Res. Commun. 247, 636-643, 1998  
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe  
A:Reference number: JEO216; PMID:98321180; PMID:9647746  
A:Accession: JEO216  
A:Molecule type: DNA  
A:Residues: 1-278 <RED>  
A:Cross-references: GB:AF062761

Query.Match 48.2%; Score 53; DB 2; Length 278;  
Best Local Similarity 61.5%; Pred. No. 0.87;  
Matches 8: Conservative 3: Mismatches 2: Indels 0: Gaps 0:

Qy	3	TVGVFGLKHDWDG	15
		:       :	
Qb	62	TVALYGLKODWEG	74

## RESULT 6

T20109  
hypothetical protein C50F4.2 - *Caenorhabditis elegans*  
C/Species: *Caenorhabditis elegans*  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #t  
C/Accession: T20109  
R/McMurray, A.  
submitted to the EMBL Data Library, Apr'il 1996  
A/Reference number: Z19225  
A/Accession: T20109  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-756 <Wtl>  
A/Cross-references: EMBL:Z70750; PIDN:CAA94737.1; GSP  
A/Experimental source: clone C50F4  
C/Genetics:  
A/Gene: CESP:C50F4.2  
A/Map position: 5  
A/Introns: 24/1; 245/3; 315/2; 399/3; 537/3; 568/2; 7  
C/Superfamily: 6-phosphofructokinase, eukaryotic type

Query Match 45.5%; Score 50; DB 2; Length 756;  
Best Local Similarity 57.1%; Pred. No. 7.8;  
Matches 8: Conservative 2; Mismatches 4; Indels

Qy 2 STVGVEGLKHDWDG 15  
| : | | : | | |  
Db 432 SGIOVIGIKHGWGD 445

## RESULT 7

JE0221  
28k surface antigen 2 - Ehrlichia canis  
C Species: Ehrlichia canis  
C Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 21-Jul-2000  
C Accession: JE0221  
R Reddy, G.R.; Sulisonga, C.R.; Barbet, A.F.; Mahan, S.M.; Burrige, M.J.; Al  
Biochem. Biophys. Res. Commun. 247, 636-643, 1998  
A Title: Molecular characterization of a 28kDa surface antigen gene family  
A Reference number: JE0216; MUID:98321180; PMID:9647746  
A Accession: JE0221  
A Molecule type: DNA  
A Residues: 1-133 <RED>  
A Cross-references: GB:AF062762; NID:G3327964; PIDN:AAC26722.1; PID:G3327964

```
Query Match      43.6%; Score 48; DB 2; Length 133;
Best Local Similarity 50.0%; Pred. No. 2.5;
Matches          9; Conservative      2; Mismatches      7; Indels      0; Gaps      0;
```

Qy 1 KSTVGVFGLKHDWDGSP 18  
Db 60 KKTTVVVGLENWAGDAI 77

## RESULT 8

R03091.8  
tauripine dehydrogenase (EC 1.5.1.23) [validated] - polychaete (*Arabella iricolor*)  
A59226  
C:Species: *Arabella iricolor*  
C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 02-Mar-2001  
C:Accession: A59226  
R:Kan-no, N.; Sato, M.; Yokoyama, T.; Nagahisa, E.  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: A59226  
A:Accession: A59226  
A:Molecule type: protein  
A:Residues: 1-396 <KAN>  
A:Experimental source: whole body  
C:Superfamily: *Arabella iricolor* tauripine dehydrogenase  
C:Keywords: oxidoreductase

Query Match 42.7%; Score 47; DB 2; Length 396;

Best Local Similarity 75.0%; Pred. No. 12;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 11 HDWDGSP1 18  
Db 221 HDWDGKPV 228

RESULT 9  
A97634  
hypothetical protein AGR\_C\_4151 [imported] - Agrobacterium tumefaciens (strain C58, Cere  
C;Species: Agrobacterium tumefaciens  
C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
C;Accession: A97634  
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, P.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A;Reference number: A97359; MUID:21608551; PMID:11743194  
A;Accession: A97634  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-385 <KUR>  
A;Cross-references: GB:AE007869; PIDN:AAK8026.1; PID:gl15157443; GSPDB:GN00169  
C;Genetics:  
A;Gene: AGR\_C\_4151  
A;Map position: circular chromosome

Query Match 41.8%; Score 46; DB 2; Length 385;  
Best Local Similarity 50.0%; Pred. No. 16;  
Matches 8; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

QY 4 VGVFGLKHDWGSFIL 19  
Db 148 LGLFGM--DWDSTPFL 161

RESULT 10  
AC2857  
conserved hypothetical protein Atu2284 [imported] - Agrobacterium tumefaciens (strain C5  
C;Species: Agrobacterium tumefaciens  
C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C;Accession: AC2857  
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I  
erge, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A;Reference number: AB2577; MUID:21608550; PMID:11743193  
A;Accession: AC2857  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-385 <KUR>  
A;Cross-references: GB:AE008689; PIDN:AAL43273.1; PID:gl17740760; GSPDB:GN00186  
C;Genetics:  
A;Gene: Atu2284  
A;Map position: circular chromosome

Query Match 41.8%; Score 46; DB 2; Length 385;  
Best Local Similarity 50.0%; Pred. No. 16;  
Matches 8; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

QY 4 VGVFGLKHDWGSFIL 19  
Db 148 LGLFGM--DWDSTPFL 161

RESULT 11  
T32146  
hypothetical protein Cl3A2.11 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T32146  
R;Rohlfing, T.; Wohldmann, P.  
submitted to the EMBL Data Library, September 1997  
A;Description: The sequence of C. elegans cosmid Cl3A2.  
A;Reference number: Z21126  
A;Accession: T32146  
A;Status: preliminary; translated from GB/EMBL/DDBU  
A;Molecule type: DNA  
A;Residues: 1-231 <ROH>  
A;Cross-references: EMBL:AF022967; PIDN:AA869881.1; GSPDB:GN00023; CESP:Cl3A2.11  
A;Experimental source: strain Bristol N2; clone Cl3A2  
C;Genetics:  
A;Gene: CESP:Cl3A2.11  
A;Map position: 5  
A;Introns: 7/3; 33/3; 71/2; 219/1

Query Match 41.4%; Score 45.5; DB 2; Length 231;  
Best Local Similarity 28.6%; Pred. No. 11;  
Matches 8; Conservative 7; Mismatches 4; Indels 9; Gaps 1;

QY 2 STVGVGFLKHDW-----DGSPILK 20  
Db 56 NSMGIYGSKRDKWKLKQTNRLTSSPLIK 83

RESULT 12  
F75010  
hypothetical protein PAB1286 - Pyrococcus abyssi (strain Orsay)  
C;Species: Pyrococcus abyssi  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C;Accession: F75010  
R;anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome str  
A;Reference number: A75001  
A;Accession: F75010  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-103 <KAW>  
A;Cross-references: GB:AJ248288; GB:AL096836; NID:95458960; PIDN:CAB50524.1; PID:954590  
A;Experimental source: strain Orsay  
C;Genetics:  
A;Gene: PAB1286  
C;Superfamily: Pyrococcus horikoshii hypothetical protein PH0346

Query Match 40.9%; Score 45; DB 2; Length 103;  
Best Local Similarity 35.0%; Pred. No. 5.6;  
Matches 7; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 KSTVGVGFLKHDWGSFILK 20  
Db 52 RTTFEIFGPKVEFSGTPYIK 71

RESULT 13  
B89834  
teichoic acid biosynthesis protein D [imported] - Staphylococcus aureus (strain N315)  
C;Species: Staphylococcus aureus  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C;Accession: B89834  
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogu  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A;Reference number: A89758; MUID:21311952; PMID:11418146  
A;Accession: B89834  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-132 <KUR>  
A;Cross-references: GB:BA000018; PID:gl3700532; PIDN:BAB41829.1; GSPDB:GN00149

	..	—
		—
	..	—
	..	—
	..	—

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2003, 13:32:19 ; Search time 22 Seconds  
(without alignments)  
42.752 Million cell updates/sec

Title: US-10-054-354-1  
Perfect score: 110  
Sequence: 1 KSTVGFGLKHDWDGSPILK 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	45.5	756	1 K6PF_CABEL	Q27483 caenorhabdi
2	46	41.8	1182	1 CGA2_HELPY	P55746 helicobacte
3	45	40.9	570	1 ILVD_LACLA	Q02139 lactococcus
4	45	40.9	704	1 DPOL_BPT3	P20311 bacterioph
5	45	40.9	704	1 DPOL_BPT7	P00581 bacterioph
6	44	40.0	160	1 MENG_DEIRA	Q9rw10 deinococcus
7	44	40.0	368	1 Y093_HAETN	P44509 haemophilus
8	43	39.1	255	1 YABD_BACSU	P27545 bacillus su
9	43	39.1	1186	1 CGA2_HELPY	P55980 helicobacte
10	42.5	38.6	84	1 N19M_HUMAN	Q95167 homo sapien
11	42	38.2	347	1 NU2M_BOVIN	P03892 bos taurus
12	42	38.2	347	1 NU2M_HIPAM	Q92220 hippopotamu
13	42	38.2	347	1 NU2M_SHEEP	O78748 ovis aries
14	42	38.2	475	1 MTHC_DROME	P83119 drosophila
15	42	38.2	3344	1 POLG_PRRVH	Q01901 p genome po
16	41	37.3	102	1 YMSA_RHIME	Q07602 thizobium m
17	41	37.3	216	1 TRPF_METKA	Q8tx29 methanopyru
18	41	37.3	414	1 KC12_RAT	Q62762 rattus norv
19	41	37.3	415	1 KC12_HUMAN	P78368 homo sapien
20	41	37.3	1081	1 AT18_HUMAN	Q8ce60 homo sapien
21	40	36.4	203	1 LIPB_BUCBP	Q89a18 buchnera ap
22	40	36.4	221	1 TRPF_HALVO	P52563 halobacteri
23	40	36.4	246	1 CVSQ_ECOLI	P22255 escherichia
24	40	36.4	326	1 GBLP_NICPL	P93340 nicotiana p
25	40	36.4	326	1 GBLP_TOBAC	P49026 nicotiana t
26	40	36.4	377	1 YAG7_METTH	O27139 methanobact
27	40	36.4	483	1 PEFD_HAETN	P44817 haemophilus
28	40	36.4	567	1 UREL_PROMI	P17086 proteus mir
29	40	36.4	578	1 NADB_ANASP	Q8yxj6 anabaena sp
30	40	36.4	600	1 YB14_MYCFN	P75448 mycoplasma
31	40	36.4	604	1 NET1_HUMAN	O95631 homo sapien
32	40	36.4	604	1 NET1_MOUSE	Q09118 mus musculu
33	40	36.4	606	1 NET1_CHICK	Q90922 gallus gall

34 40 36.4 792 1 OSTA\_XYLFA  
35 40 36.4 885 1 PLSE\_XANAC  
36 40 36.4 1520 1 PMPD\_CHLMU  
37 39.5 35.9 419 1 CHSD\_PETHY  
38 39.5 35.9 518 1 NAEK\_BRUME  
39 39.5 35.9 525 1 NAEK\_BRUSU  
40 39.5 35.9 952 1 KP58\_DROME  
41 39 35.5 92 1 PAK2\_ANASP  
42 39 35.5 113 1 SU11\_SALBA  
43 39 35.5 226 1 TRF2\_METMA  
44 39 35.5 284 1 YOAK\_BACSU  
45 39 35.5 371 1 Y028\_ARCFU

## ALIGNMENTS

RESULT 1  
K6PF\_CABEL  
ID K6PF\_CABEL STANDARD; PRT; 756 AA.  
AC Q27483;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Probable 6-phosphofructokinase (EC 2.7.1.11) (Phosphofructokinase)  
DE (Phosphohexokinase).  
GN CS0F4.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
ON NCBI\_TaxID=6239;  
RX [1]  
RC SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA McMurray A.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
CC -|- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-fructose 1,6-bisphosphate.  
CC -|- PATHWAY: Key control step of glycolysis.  
CC -|- SIMILARITY: Belongs to the phosphofructokinase family. Two domains subfamily.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
-----

EMBL; Z70750; CAA94737.1; -.  
PIR; T20109; T20109.  
HSP; P00512; 3PFK.  
WormRep; CS0F4.2; CE05467.  
DR InterPro; IPR000023; Pffruckinase.  
DR Pfam; PF00365; PFK; 2.  
DR PRINTS; PR00476; PHFRCTKINASE.  
DR ProDom; PD000707; Pffruckinase; 2.  
DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 1.  
DR Kinase; Transfrase; Glycolysis; Repeat.  
SQ SEQUENCE 756 AA; 83301 MW; 26A89B801D286534 CRC64;

Query Match 45.5%; Score 50; DB 1; Length 756;  
Best Local Similarity 57.1%; Pred. No. 2.7;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 STVGFGLKHDWDG 15  
Db 432 SGIVIGIKHGDG.445

RESULT 2  
CGA2\_HELPY



```

CC      + {DNA} (N).
CC      -!- SUBUNIT: COMPOSED OF TWO SUBUNITS. ONE IS ENCODED BY THE PHAGE AND
CC      THE OTHER IS ENCODED BY THE HOST THIOREDOXIN.
CC      -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-A FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; X17255; CAA35140.1; --
CC      PIR; S07512; S07512.
CC      HSP; P00581; 1T7P.
CC      InterPro; IPR001098; DNA_pol.
CC      InterPro; IPR002298; DNA_pol.
CC      Pfam; PF00476; DNA_pol_A; 1.
CC      PRINTS; PR00868; DNAPOLI.
CC      SMART; SM00482; POLAC; 1.
CC      PROSITE; PS00447; DNA_POLYMERASE A; 1.
CC      Transferase; DNA-directed DNA polymerase; DNA replication;
CC      DNA-binding; Hydrolase; Exonuclease
CC      SEQUENCE 704 AA; 79985 MW; D65DFED99AE31234 CRC64;
CC
CC      Query Match 40.9%; Score 45; DB 1; Length 704;
CC      Best Local Similarity 87.5%; Pred. No. 16;
CC      Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
CC
CC      QY 8 GLKHDWDG 15
CC      |||||
CC      Db 637 GLKHGWDG 644

```

```

RESULT 5
DPOL_BPT7
ID DPOL_BPT7 STANDARD; PRT; 704 AA.
AC P00581;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase (EC 2.7.7.7) (T7 DNA polymerase).
DE 5.
OS Bacteriophage T7.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like viruses.
OC NCBI_TaxID=10760;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=83241725; PubMed=6864790;
RX Dunn J.J., Studier F.W.;
RA "Complete nucleotide sequence of bacteriophage T7 DNA and the
RT locations of T7 genetic elements.";
RL J. Mol. Biol. 166:477-535(1983).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=84164887; PubMed=6708104;
RA Moffatt B.A., Dunn J.J., Studier F.W.;
RT "Nucleotide sequence of the gene for bacteriophage T7 RNA
RT polymerase.";
RL J. Mol. Biol. 173:265-269(1984).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX MEDLINE=98101638; PubMed=9440588;
RA Double S., Tabor S., Long A.M., Richardson C.C., Ellenberger T.;
RT "Crystal structure of a bacteriophage T7 DNA replication complex at
RT 2.2-A resolution.";
RL Nature 391:251-258(1998).
CC -!- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
CC EXHIBITS A 3' TO 5' EXONUCLEASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA} (N).

```

```

CC      -!- SUBUNIT: COMPOSED OF TWO SUBUNITS. ONE IS ENCODED BY THE PHAGE AND
CC      THE OTHER IS ENCODED BY THE HOST THIOREDOXIN.
CC      -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-A FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; V01146; CAA24412.1; --
CC      PIR; A00716; DJBPT7.
CC      PDB; 1T7P; 25-FEB-98.
CC      InterPro; IPR001098; DNA_pol.
CC      InterPro; IPR002298; DNA_pol.
CC      Pfam; PF00476; DNA_pol_A; 1.
CC      PRINTS; PR00868; DNAPOLI.
CC      SMART; SM00482; POLAC; 1.
CC      PROSITE; PS00447; DNA_POLYMERASE A; 1.
CC      Transferase; DNA-directed DNA polymerase; DNA replication; Hydrolase;
CC      Exonuclease; DNA-binding; 3D-structure.
CC      SEQUENCE 704 AA; 79691 MW; 17089CE2AD9FB596 CRC64;
CC
CC      Query Match 40.9%; Score 45; DB 1; Length 704;
CC      Best Local Similarity 87.5%; Pred. No. 16;
CC      Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
CC
CC      QY 8 GLKHDWDG 15
CC      |||||
CC      Db 637 GLKHGWDG 644

```

```

RESULT 6
MENG_DEIRA
ID MENG_DEIRA STANDARD; PRT; 160 AA.
AC Q9RW10;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable S-adenosylmethionine:2-demethylmenaquinone methyltransferase
DE (EC 2.1.1.-.-).
DE MENG OR DR0859.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
CC -!- FUNCTION: CONVERTS DIMETHYLMENAQUINONE (DMK) TO MENAQUINONE (MK)
CC (By similarity).
CC -!- PATHWAY: Menaquinone biosynthesis; last step.
CC -!- SIMILARITY: BELONGS TO THE MENG FAMILY.
CC -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC EMBL; AE001940; AAF10437.1; --  
 DR PIR; A75466; A75466.  
 DR TIGR; DR0859; --  
 DR HAWAP; MF 00471; --; 1.  
 DR InterPro; IPR005493; Methyltransf 6.  
 DR Pfam; PF03737; Methyltransf 6; 1\_--  
 KW Menquinone biosynthesis; Transferase; Methyltransferase;  
 KW Complete proteome.  
 SQ SEQUENCE 160 AA; 16860 MW; 54DF226CA19EC0E CRC64;

Query Match 40.0%; Score 44; DB 1; Length 160;  
 Best Local Similarity 50.0%; Pred. No. 5.1;  
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 4 VGVFLKHDWDG 15  
 :|||:::|  
 Db 79 LGVFGVNGWEG 90

# RESULT 7

Y093 HAEIN STANDARD; PRT; 368 AA.  
 AC P44509;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein HI0093.  
 GN HI0093.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellales; Haemophilus.  
 OC NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kierliavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geobhagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";  
 RL Science 269:496-512(1995).  
 CC -1- SIMILARITY: BELONGS TO THE CDAR FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U32695; AAC21771.1; --  
 DR PIR; E64142; E64142.  
 DR TIGR; HI0093; --  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 368 AA; 42251 MW; C77F1C9EF043B89A CRC64;

Query Match 40.0%; Score 44; DB 1; Length 368;  
 Best Local Similarity 46.2%; Pred. No. 12;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 8 GLKHDWDGSPILK 20  
 :|||:::|  
 Db 283 GLSHWQGNELIK 295

## RESULT 8

YABD\_BACSU STANDARD; PRT; 255 AA.  
 AC P37545;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putative deoxyribonuclease yabd (EC 3.1.21.-).  
 GN YABD.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OC NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=96051385; PubMed=7584024;  
 RA Ogasawara N., Nakai S., Yoshikawa H.;  
 RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
 RT subtilis chromosome containing the replication origin";  
 RL DNA Res. 1:1-14(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Etian K.D., Erington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaeser P., Goifeau A., Gollightly E.J., Grandi G.,  
 RA Giuseppe G., Guy B.J., Haga K., Haeche J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaert-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Maesuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale F.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Takatoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tazato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RT subtilis";  
 RL Nature 390:249-256(1997).  
 CC -1- SIMILARITY: BELONGS TO THE TATD DNASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; D26185; BAA05274.1; --  
 DR EMBL; Z99104; CAB11815.1; --  
 DR PIR; S66069; S66069.  
 DR Subtilist; BG10102; yabd.  
 DR InterPro; IPR001130; TatD\_DNase.  
 DR Pfam; PF01026; TatD\_DNase; 1.



DR TIGR00010; TIGR00010; 1.  
 DR PROSITE; PS01137; TATD 1; 1.  
 DR PROSITE; PS01090; TATD 2; 1.  
 DR PROSITE; PS01091; TATD 3; 1.  
 KW Hypothetical protein; Hydrolase; Nuclease; Complete proteome.  
 SQ SEQUENCE 255 AA; 29232 MW; ADF8EE9F5AEB0ED4 CRC64;

Query Match 39.1%; Score 43; DB 1; Length 255;  
 Best Local Similarity 50.0%; Pred. No. 12;  
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 VGVFGLKHDWDGSP 17  
 : ||| :  
 Db 90 IGEWGLDIHWKSP 103

## RESULT 9

CAGA\_HELPY STANDARD; PRT; 1186 AA.

AC P55980;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cytotoxicity associated immunodominant antigen (120 kDa protein)  
 DE (CAG pathogenicity island protein 26).  
 GN CAGA OR CAI OR CAG26 OR HP0547.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
 OC Helicobacteraceae; Helicobacter.  
 OX NCBI\_TaxID=210;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=26695 / ATCC 700392;  
 RX MEDLINE=97394467; PubMed=9252185;  
 RA Tomb J.-F., White O., Kervilave A.R., Clayton R.A., Sutton G.G.,  
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,  
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,  
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,  
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the gastric pathogen Helicobacter  
 RT pylori."  
 RL Nature 388:539-547(1997).  
 CC -!- FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT,  
 CC OR FUNCTION OF THE CYTOTOXIN.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AE000569; AAD07614.1; -  
 DR TIGR; HP0547; -  
 DR InterPro; IPR005169; Caga.

DR InterPro; IPR004355; IVSec\_caga.  
 DR Pfam; PF03507; Caga; 1.  
 DR PRINTS; PR01553; TYPE4SSCAGA.

KW Antigen; Complete proteome.  
 FT DOMAIN 247; 250 POLY-THR.  
 FT DOMAIN 883 889 POLY-ASN.

SQ SEQUENCE 1186 AA; 132386 MW; B05C3F2CCC4444F4 CRC64;

Query Match 39.1%; Score 43; DB 1; Length 1186;  
 Best Local Similarity 58.3%; Pred. No. 60;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 VGVFGLKHDWDGS 16  
 : ||| :  
 Db 328 GGFGLKHDWNAT 339

## RESULT 10

NI9M\_HUMAN STANDARD; PRT; 84 AA.

AC O95167;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE NADH-ubiquinone oxidoreductase B9 subunit (EC 1.6.5.3) (EC 1.6.99.3)  
 DE (Complex I-B9) (CI-B9).  
 GN NDUFAB3.  
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99097250; PubMed=9878551;  
 RA Loeffen J.L.C.M., Triepels R.H., van den Heuvel L., Schuelke M.,  
 RA Buskens C.A.F., Smeets R.J.P., Trijbels J.M.P., Smeitink J.A.M.;  
 RT "cDNA of eight nuclear encoded subunits of NADH:ubiquinone  
 RT oxidoreductase: human complex I cDNA characterization completed."  
 RL Biochem. Biophys. Res. Commun. 253:415-422(1998).  
 RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Blood;  
 RX MEDLINE=20499367; PubMed=11042152;

RA Zhang Q.-H., Ye M., Wu X.-Y., Ren S.-X., Zhao M., Zhao C.-J., Fu G.,  
 RA Shen Y., Fan H.-Y., Lu G., Zhong M., Xu X.-R., Han Z.-G., Zhang J.-W.,  
 RA Tao J., Huang Q.-H., Zhou J., Hu G.-X., Gu J., Chen S.-J., Chen Z.;  
 RT "Cloning and functional analysis of cDNAs with open reading frames for  
 RT 300 previously undefined genes expressed in CD34+ hematopoietic  
 RT stem/progenitor cells."  
 RL Genome Res. 10:1546-1560(2000).  
 RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE=Prostate;  
 RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullahy P.H.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Munz D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences."

CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -!- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY  
 CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED  
 CC TO BE UBIQUINONE.

CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.

CC -!- CATALYTIC ACTIVITY: NADH + acceptor = NAD(+) + reduced acceptor.

CC -!- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 40 DIFFERENT SUBUNITS.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
 CC inner membrane.

CC -!- SIMILARITY: BELONGS TO THE COMPLEX I B9 SUBUNIT FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

[illegible]

ID	NU2M SHEEP	STANDARD;	PRT;	347 AA.
AC	O78748;			

[illegible]

DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE NADH-ubiquinone oxidoreductase chain 2 [EC 1.6.5.3].  
 GN MTND2 OR ND2.  
 OS Ovis aries (Sheep).  
 OC Mitochondrion.  
 OG Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OO Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OT Bovidae; Caprinae; Ovis.  
 NCBI\_TaxID=9940;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=Wernholandschaf; TISSUE=Liver;  
 RX MEDLINE=98404761; PubMed=9767689;  
 RT Hiedler S., Lewalski H., Wassmuth R., Janke A.;  
 RR "The complete mitochondrial DNA sequence of the domestic sheep (*Ovis*  
 RL aries) and comparison with the other major ovine haplotype." ;  
 RM J. Mol. Evol. 47:441-448(1998).  
 CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
 CC inner membrane.  
 CC -!- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 2 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC at European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to licensee@isb-sib.ch).  
 CC -----  
 EMBL AF010406; AACD10097.1; -.  
 DR PIR; T11051; T11051.  
 DD InterPro: IPR003917; NADHub\_oxred42.  
 DR InterPro: IPR001750; Oxidored\_q1.  
 DR Pfam; PF00361; oxigored\_q1; 1.  
 DR PRINTS; PR01436; NADHDGNAE2.  
 DK Oxidoreductases; NAD; Ubiquinone; Mitochondrion; Transmembrane.  
 KW SEQUENCE 347 AA; 39128 MW; 5202DF0A0D6C991B CRC64;  
 SQ  
 Query Match 38.2%; Score 42; DB 1; Length 347;  
 Best Local Similarity 38.9%; Pred. No. 24;  
 Matches 7; Conservative 3; Mismatches 8; Indels . 0; Gaps 0;

---

QY		2 STVGVEGLKHWDPSPIL 19	
DDB		: : :	
		223 STTTLSHTWTKAPIM 240	

---

```

RESULT 14
MTHC DROME
IID_MTHC_DROME STANDARD; PRT; 475 AA.
AAC_P83119;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DR Probable G-protein-coupled receptor Mth-like 12 precursor (Methuselaha- like 12 protein)
DN MTHL12 OR MTH-LIKE-12.
OS Drosophila melanogaster (Fruit fly).
OG Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephrydoidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F., George K.R., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

```

FT	DOMAIN	264	275	EXTRACELLULAR (POTENTIAL).	CC	TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.
FT	TRANSSEM	276	296	3 (POTENTIAL).	CC	FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
FT	DOMAIN	297	307	CYTOPLASMIC (POTENTIAL).	CC	MAY BE INVOLVED IN REPLICATION
FT	TRANSSEM	308	328	4 (POTENTIAL).	CC	FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.
FT	DOMAIN	329	360	EXTRACELLULAR (POTENTIAL).	CC	CATALYTIC ACTIVITY: Hydrolyzes glutamyl bonds, and activity is
FT	TRANSSEM	361	381	5 (POTENTIAL).	CC	further restricted by preferences for the amino acids in P6 - P1,
FT	DOMAIN	382	403	CYTOPLASMIC (POTENTIAL).	CC	that vary with the species of potyvirus, e.g. Glu-Xaa-Xaa-Tyr-Xaa-
FT	TRANSSEM	404	424	6 (POTENTIAL).	CC	Gln+(Ser or Gly) for the enzyme from tobacco etch virus. The
FT	DOMAIN	425	442	EXTRACELLULAR (POTENTIAL).	CC	natural substrate is the viral polyprotein, but other proteins and
FT	TRANSSEM	443	463	7 (POTENTIAL).	CC	oligopeptides containing the appropriate consensus sequence are
FT	DOMAIN	464	475	CYTOPLASMIC (POTENTIAL).	CC	also cleaved.
FT	DISULFID	27	81	BY SIMILARITY.	CC	-1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
FT	DISULFID	83	88	BY SIMILARITY.	CC	{RNA} (N).
FT	DISULFID	92	183	BY SIMILARITY.	CC	-1- CATALYTIC ACTIVITY: Hydrolyzes a Gly- -Gly bond at its own C-
FT	DISULFID	93	104	BY SIMILARITY.	CC	terminus, commonly in the sequence -Tyr-Xaa-Val-Gly- -Gly, in the
FT	DISULFID	149	203	BY SIMILARITY.	CC	processing of the potyviral polyprotein.
FT	CARBOHYD	19	19	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	-1- PTM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.
FT	CARBOHYD	34	34	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	-1- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE
FT	CARBOHYD	55	55	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
FT	CARBOHYD	135	135	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
FT	CARBOHYD	352	352	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	INDIVIDUAL PROTEINS.
FT	SEQUENCE	475 AA; 55424 MW; 3590FEF875F77264 CRC64;			CC	-1- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
Qy	4 VGVFGLKHDWGS 16				CC	-1- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
Db	342 VGVFGLFTDWNSS 354				CC	-1- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.
RESULT 15					CC	-----
POLG PRSVH					CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
ID	POLG PRSVH	STANDARD;	PRT; 3344 AA.		CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
AC	Q01901;				CC	the European Bioinformatics Institute. There are no restrictions on its
DT	01-OCT-1993 (Rel. 27, Created)				CC	use by non-profit institutions as long as its content is in no way
DT	01-OCT-1996 (Rel. 34, Last sequence update)				CC	modified and this statement is not removed. Usage by and for commercial
DT	28-FEB-2003 (Rel. 41, Last annotation update)				CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
DE	Genome polyprotein [Contains: N-terminal protein (P1); Helper				CC	or send an email to license@sib-sib.ch).
DE	component proteinase (EC 3.4.22.45) (HC-Pro); Protein P3; 6 kDa				CC	-----
DE	protein 1 (6k1); Cytoplasmic inclusion protein (CI); 6 kDa protein 2				CC	EMBL; X67673; CAA47905.1; -;
DE	(6k2); Genome-linked protein (VPG); Nuclear inclusion protein A (NI-A)				CC	EMBL; X67672; AAB23789.1; -;
DE	(NTA) (EC 3.4.22.44) (49 kDa-proteinase) (49 kDa-Pro); Nuclear				CC	PIR; JQ1899; JQ1899.
DE	inclusion protein B (NT-B) (NIB) (RNA-directed RNA polymerase)				CC	MEROPS; C04.009; -;
DE	(EC 2.7.7.48); Coat protein (CP).)				CC	MEROPS; C06.001; -;
OS	Papaya ringspot virus (strain P / mutant HA).				CC	MEROPS; S30.001; -;
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;				CC	InterPro; IPR001410; DEAD.
OC	Potyvirus.				CC	InterPro; IPR001650; Helicase_C.
OX	NCHI_TaxID=31731;				CC	InterPro; IPR001730; Peptidase_C4.
RP	SEQUENCE FROM N.A.				CC	InterPro; IPR001456; Peptidase_C6.
RA	Wang C.H., Bau H.J., Yeh S.D.;				CC	InterPro; IPR001592; Poty_coat.
RT	"Comparison of the nuclear inclusion b protein and coat protein genes				CC	InterPro; IPR002540; Poty_P1.
RT	of five papaya ringspot virus strains distinct in geographic origin				CC	InterPro; IPR007095; RNA_pol_DS_PS.
RT	and pathogenicity."				CC	InterPro; IPR001205; RNA_pol_P3D.
RL	Phytopathology 84:1205-1210 (1994).				CC	InterPro; IPR007094; RNA_pol_PSVir.
RP	SEQUENCE FROM N.A.				CC	Pfam; PF00271; helicase_C1.
RX	MEDLINE=93019006; PubMed=1402799;				CC	Pfam; PF00863; Peptidase_C4; 1.
RA	Yeh S.D., Jan F.J., Chiang C.H., Doong T.J., Chen M.C.,				CC	Pfam; PF00851; Peptidase_C6; 1.
RA	Chung P.H., Bau H.J.;				CC	Pfam; PF00767; Poty_coat; 1.
RT	"Complete nucleotide sequence and genetic organization of papaya				CC	Pfam; PF01577; Poty_P1; 1.
RT	ringspot virus RNA."				CC	Pfam; PF00680; RNA_dep_RNA_pol; 1.
RL	J. Gen. Virol. 73:2531-2541 (1992).				CC	PRINTS; PR00966; NTAPOPTYASE.
RL	[3]				CC	SMART; SM00487; DEXDC; 1.
RP	SEQUENCE OF 2561-3344 FROM N.A.				CC	SMART; SM00490; HELIC; 1.
RX	MEDLINE=93090098; PubMed=1456896;				CC	KW Hydrolase; Transferase; Thiol protease; RNA-directed RNA polymerase;
RA	Wang C.H., Yeh S.D.;				CC	KW Coat protein; Polyprotein; Covalent protein-RNA linkage; Helicase;
RT	"Nucleotide sequence comparison of the 3'-terminal regions of severe,				CC	KW ATP-binding.
RT	RT mild, and non-papaya infecting strains of papaya ringspot virus."				CC	FT CHAIN 1 529 N-TERMINAL PROTEIN (BY SIMILARITY).
RL	Arch. Virol. 127:345-354 (1992).				CC	FT CHAIN 530 1149 HELPER COMPONENT PROTEINASE
CC	-1- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID				CC	FT CHAIN 1150 ? PROTEIN P3 (BY SIMILARITY).
					CC	FT CHAIN ? 1401 6 kDa PROTEIN 1 (BY SIMILARITY).
					CC	FT CHAIN 1402 2036 CYTOPLASMIC INCLUSION PROTEIN (BY
					CC	FT CHAIN 2037 2093 6 kDa PROTEIN 2 (BY SIMILARITY).
					CC	FT CHAIN 2094 ? GENOME-LINKED PROTEIN (BY SIMILARITY).
					CC	FT CHAIN ? 2520 NUCLEAR INCLUSION PROTEIN A
					CC	FT CHAIN 2521 3037 (BY SIMILARITY).
					CC	FT CHAIN 2521 3037 NUCLEAR INCLUSION PROTEIN B

FT CHAIN 3038 3344 (BY SIMILARITY).  
FT BINDING 2156 2156 COAT PROTEIN (BY SIMILARITY).  
FT NP BIND 1486 1493 COVALENT LINKAGE OF VIRAL RNA (BY  
SQ SEQUENCE 3344 AA; 381040 MW; E90CD7523AC5243D CRC64;  
SIMILARITY).  
ATP (POTENTIAL).

Query Match 38.2%; Score 42; DB 1; Length 3344;  
Best Local Similarity 61.5%; Pred. No. 2.6e+02;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 8 GLKHDWDGSPILK 20  
||| ||| :|  
Db 2049 GLKGVWDGSLMIK 2061

Search completed: September 5, 2003, 14:07:07  
Job time : 23 secs

